

Appl. No. 09/885,356
Supp. Amdt. Dated May 21, 2004
Reply to Examiner Interview of May 19, 2004

Amendments to the Specification:

Please replace previously amended paragraph [0006] with the following amended paragraph:

[0006] According to one embodiment of the present invention, a method of optimizing a response time for retrieving relevant documents from a set of candidate documents is provided. The candidate documents are identified in response to a search query where the search query includes one or more terms. A term weight is assigned to each of the terms in the search query. Documents are associated to a relevance score bin based on a total matched term weight that is based on a sum of the term weights of the matched terms in the search query, where a document that matches a first total term weight is associated to a more relevant score bin than a document that matches a second total term weight less than the first total term weight. A set of most relevant documents are then retrieved based on the association to the relevance score bins having a highest relevance score without retrieving other candidate documents.

Please replace previously amended paragraph [0007] with the following amended paragraph:

[0007] According to another embodiment of the present invention, an information retrieval system is provided. The system includes logic for processing a search query that has one or more terms. A document retrieval logic identifies candidate documents that match the search query. A ranking logic assigns a term weight to each of the terms of the search query and associates each combination of matched query term weights to a relevance score range. The ranking logic also groups the candidate documents based on the total matched term weight, which is based on a sum of the term weights of the matched terms in the search query, where a document that matches a first total term weight is associated to a more relevant score range than a document that matches a second total term weight that is less than the first total term weight. A retrieval logic then retrieves a set of relevant documents associated to the relevance score ranges having a greatest matched term weight without receiving the candidate documents from other relevance score ranges.

Appl. No. 09/885,356
 Supp. Amdt. Dated May 21, 2004
 Reply to Examiner Interview of May 19, 2004

Please replace the previously amended paragraph [0026] with the following amended paragraph:

[0026] A final relevance score for a document is a function of the term relevance score and the term weight for each query term the document matches. However, before the final relevance score is computed, the ranking logic 135 establishes a relationship between the total term weight matched by a document, which is based on a sum of the term weights of the terms in the query that are matched by the document, and a range of final relevance scores that can be assigned to the document. Thus, the total matched term weight computed for a document is a primary factor for determining the final relevance score of the document. In this manner, a document that matches a first total term weight from the query will always have a greater relevance score than a document that matches a second total term weight that is less than the first total term weight. This is described in greater detail below. As a result of this relationship, the most relevant documents can be identified without having to compute the relevance scores of the entire set of candidate documents. Thus, a small set of most relevant documents can be retrieved without having to retrieve the entire set of candidate documents thereby increasing response time.

Please replace paragraph [0029] with the following amended paragraph:

[0029] Once the terms from the query are identified, the system identifies documents that match one or more of the terms from the query or their themes (block 210) and builds a hitlist for each term. The hitlist can be a table, index, tree or the like that identifies which documents matched each term. Suppose that the selectivity of the terms are as follows in Table (1):

Table (1):

Term	Document Hits
Oracle Text:	1,000
Adoption:	50,000
Japan:	50,000,000

Please replace paragraph [0030] with the following amended paragraph:

Appl. No. 09/885,356
Supp. Amdt. Dated May 21, 2004
Reply to Examiner Interview of May 19, 2004

[0030] Each term from the query is assigned a term weight based on its occurrence frequency (block 215). Inverse frequency scoring is one algorithm that can be used. Terms that occur infrequently are assigned a greater term weight than terms that ~~are~~ occur frequently. For this example, the term weights are assigned as 4 for "oracle text", 2 for "adoption", and 1 for "Japan."